

# Molecular Insight from DFT Computations and Kinetic Measurements into the Steric Factors Influencing Peptide Bond Hydrolysis Catalyzed by a Dimeric Zr(IV)-Substituted Keggin Type Polyoxometalate

Tzvetan T. Mihaylov,<sup>\*,†</sup> Hong Giang T. Ly,<sup>‡</sup> Kristine Pierloot<sup>†</sup> and Tatjana N. Parac-Vogt<sup>‡</sup>

<sup>†</sup> Laboratory of Computational Coordination Chemistry, Department of Chemistry, KU Leuven, Celestijnenlaan 200F, 3001 Leuven, Belgium.

<sup>‡</sup> Laboratory of Bioinorganic Chemistry, Department of Chemistry, KU Leuven, Celestijnenlaan 200F, 3001 Leuven, Belgium.

**KEYWORDS:** Polyoxometalates / Peptide bond hydrolysis / Zirconium(IV) / Artificial peptidases / DFT calculations

## ABSTRACT

Peptide bond hydrolysis of several peptides with Gly-X sequence, (X = Gly, Ala, Val, Leu, Ile, Phe) catalyzed by a dimeric Zr(IV)-substituted Keggin type polyoxometalate (POM)  $(\text{Et}_2\text{NH}_2)_8[\{\alpha\text{-PW}_{11}\text{O}_{39}\text{Zr}(\mu\text{-OH})(\text{H}_2\text{O})\}_2]\cdot 7\text{H}_2\text{O}$  (**1**) was studied by means of kinetic experiments and  $^1\text{H}$  NMR spectroscopy. The observed rate of peptide bond hydrolysis was found to decrease with increase of the side chain bulkiness, from  $4.44\times 10^{-7}\text{ s}^{-1}$  for Gly-Gly to  $0.81\times 10^{-7}\text{ s}^{-1}$  for Gly-Ile. A thorough DFT investigation was performed to elucidate (a) the nature of the hydrolytically active species in solution, (b) the mechanism of peptide bond hydrolysis and (c) the influence of the aliphatic residues on the rate of hydrolysis. Formation of substrate-catalyst complexes of the dimeric POM, **1**, was predicted as thermodynamically unlikely. Instead, the substrates prefer to bind to the monomerization product of **1**,  $[\alpha\text{-PW}_{11}\text{O}_{39}\text{Zr}(\text{OH})(\text{H}_2\text{O})]^{4-}$  (**2**), which is also present in solution. In the hydrolytically active complex two dipeptide ligands are coordinated to the Zr(IV) center of **2**. The first ligand is bidentate bound through its amino nitrogen and amide oxygen atoms while the second one is monodentate bound through a carboxylic oxygen atom. The mechanism of hydrolysis involves nucleophilic attack by a solvent water molecule on the amide carbon atom of the bidentate bound ligand. In this process the uncoordinated carboxylic group of the same ligand acts as a general base to abstract a proton from the attacking water molecule. The decrease of the hydrolysis rate with increase of the side chain bulkiness is mostly due to the increased ligand conformational strain in the rate-limiting transition state, which elevate the reaction activation energy. The conformational strain increases first, upon substitution of  $\text{H}_\alpha$  in Gly-Gly with aliphatic  $\alpha$ -substituent and second, with the  $\beta$ -branching of the  $\alpha$ -substituent.

## INTRODUCTION

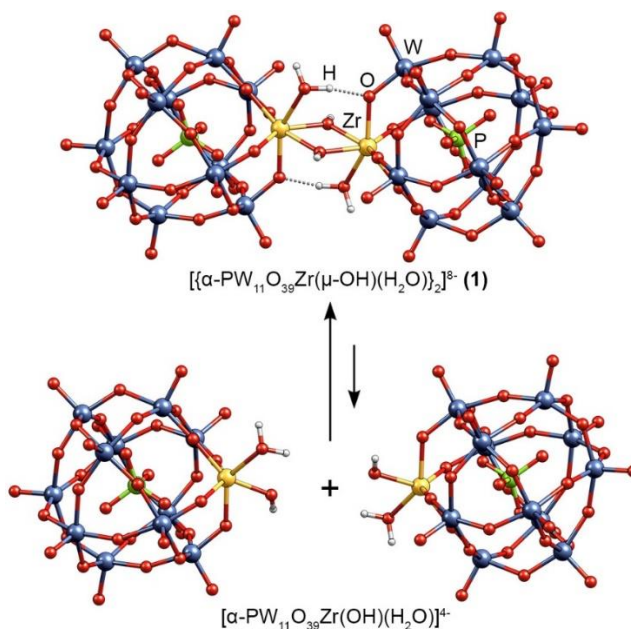
Selective hydrolysis of peptide bonds in proteins plays an important role in a broad range of research fields such as chemical biology, proteomics, and drug discovery.<sup>1</sup> The reaction is also frequently applied in structural determination of peptides and proteins. Proteolytic enzymes with a high catalytic power can be used for this purpose. However, they often operate only in a narrow range of temperature and pH. In addition, the high cost and limited or fixed selectivity of proteolytic enzymes has led to the search for alternative cleavage agents for hydrolysis of peptide bond, which is extremely stable at physiological pH and room temperature.<sup>2</sup>

Polyoxometalates (POMs) are generally described as clusters of early transition metal ions in their highest oxidation state bridged by oxo anions. Owing to fascinating properties such as tunable acidity and redox properties, high thermal stability, hydrolytic and oxidative stability, POMs have been applied in broad fields including catalysis,<sup>3</sup> materials,<sup>4</sup> and medicine.<sup>5</sup> In recent years applications of POMs for catalytic purposes have also been extensively investigated. Series of reviews and publications focusing on the use of POMs in heterogeneous catalysis,<sup>6</sup> electrocatalysis,<sup>7</sup> water oxidation,<sup>8</sup> and oxidation of organic compounds<sup>9</sup> have been published. In order to improve or tune the catalytic properties of POMs, lanthanide or transition metal ions can be introduced into the skeleton of vacant POMs to form metal-substituted POMs (MSPs). The combination of the POM framework, with its negative charge playing a role in electrostatic interaction with positively charged substrates, and the active metal sites, which can coordinate water or other ligands at their free coordination sites, makes such MSPs attractive for catalytic applications. In addition, the introduction of metal ions into POM framework can stabilize metal ions, preventing gel formation in aqueous solution.<sup>10</sup> The first investigations concerned the catalytic properties of Ti(IV) and Zr(IV)-substituted POMs towards oxidation reactions with H<sub>2</sub>O<sub>2</sub>.<sup>11</sup> The use of MSPs as catalysts for a broad range of chemical reactions, such as H<sub>2</sub>O<sub>2</sub>-based oxidation of organic compounds,<sup>12</sup> cyclization of citronellal derivatives,<sup>13</sup> Diels-Alder reactions and aldol reactions of imines,<sup>14</sup> oxygenation of catechols,<sup>15</sup> synthesis of cyclic carbonates,<sup>16</sup> oximation of aldehydes and ketones,<sup>17</sup> and the splitting of water<sup>18</sup> has recently been reported.

In our quest to develop MSPs as potential catalysts for peptide bond hydrolysis, Zr(IV)- and Ce(IV)-substituted POMs have been used, because both Zr(IV) and Ce(IV) have a high Lewis

acidity, oxophilicity, and kinetic lability, and tend to form complexes with high coordination numbers. The hydrolytic activity of a series of Zr(IV)-substituted Keggin, Lindqvist and Wells-Dawson POMs towards peptide bond hydrolysis in dipeptides<sup>10a, b, 19</sup> and oligopeptides<sup>20</sup> was discovered previously. All complexes act as homogeneous catalysts and significantly accelerate the hydrolysis reaction rates in mildly acidic and neutral pH environments. In addition, we recently showed that a Ce(IV)-substituted Keggin POM and Zr(IV)-substituted Keggin, Lindqvist and Wells-Dawson POMs selectively hydrolyzes proteins such as hen-egg-white lysozyme, albumins, and horse heart myoglobin, thus demonstrating the potential of Ce(IV)- and Zr(IV)-substituted POMs as novel artificial proteases.<sup>10c, 21</sup>

In aqueous solution, dimeric Zr(IV)-substituted POMs can partly dissociate into monomeric species, which are presumed to be the active species during the catalytic reactions.<sup>10b, 19a</sup> For example, 1:1 Zr(IV)-Keggin POM,  $[\alpha\text{-PW}_{11}\text{O}_{39}\text{Zr}(\text{OH})(\text{H}_2\text{O})]^{4-}$  was identified as the active compound for the hydrolysis of DNA-model phosphoesters both by experimental and theoretical means,<sup>22</sup> even though the dimeric 2:2 Zr(IV)-Keggin POM,  $[\{\alpha\text{-PW}_{11}\text{O}_{39}\text{Zr}(\mu\text{-OH})(\text{H}_2\text{O})\}_2]^{8-}$  (**1**), was used as a starting catalyst (see Figure 1).<sup>22</sup> This may be the reason for the superior hydrolytic activity of **1** towards myoglobin, as compared to other POMs which do not undergo dissociation.<sup>10c, 23</sup>



**Figure 1.** Molecular structures of 2:2 Zr(IV)-Keggin and 1:1 Zr(IV)-Keggin POMs.

The selectivity of **1** towards dipeptides containing different amino acid sequences (other than those studied here) was recently reported and the results indicated that the reaction rate is dependent on the nature and the bulkiness of the amino acid side chain.<sup>19c</sup> The first mechanistic study on peptide bond hydrolysis catalyzed by a dimeric tetrazirconium(IV) Wells-Dowson POM was also reported recently by us.<sup>23</sup> For dipeptides with aliphatic residues the hydrolytically active substrate-catalyst complex is formed through bidentate chelate coordination of the ligand to a Zr(IV) center of the POM, involving the amino nitrogen and the amide oxygen atoms, while the terminal ligand carboxylic group remains uncoordinated. In the rate-determining reaction step the latter acts as a general base to facilitate nucleophilic attack by a solvent water molecule on the amide carbon atom. Similar to **1**, the rates of hydrolysis within a series of dipeptides were also with this catalyst found to be dependent both on the nature and the bulkiness of the side chains. The molecular factors governing this selectivity, however, remain unclear. The large variety of amino acid sequences involved in the dipeptide series presupposes a variety of different substrate-catalyst complexes and possible mechanistic scenarios for hydrolysis. For instance, some of the dipeptides involve amino acids with residues that could coordinate to Zr(IV), such as Gly-Asn, Gly-Asp, Gly-Gln, Gly-His, Gly-Ser and Gly-Thr, while others with positively charged residues, like Gly-Lys and Gly-Arg, may stabilize the complex through electrostatic interactions with the negatively charged POM surface. Therefore, the coordination behavior and the mechanisms of hydrolysis should be clarified individually in every particular case.

This study is focused on the hydrolytic activity of **1** towards a series of dipeptides with a common Gly-X sequence, where X is an amino acid with an aliphatic (non-coordinating) side chain, such as Ala, Val, Leu, Ile and Phe. In this series both the coordination behavior and the mechanisms of hydrolysis are expected to be rather similar, which allows us to distinguish the role of the aliphatic side chain in the process. However, the tendency of the parent compound **1** for partial dissociation in solution brings additional complexity to the problem studied. The kinetics of the Gly-X dipeptides hydrolysis in presence of **1** was studied experimentally with <sup>1</sup>H NMR method. To explain the experimental findings, the nature of the hydrolytically active species, the mechanism of peptide bond hydrolysis as well as the effect of the aliphatic residues on the rate of hydrolysis were investigated on a molecular level by means of density functional methods.

## EXPERIMENTAL SECTION

### Chemicals

(Et<sub>2</sub>NH<sub>2</sub>)<sub>8</sub>[{ $\alpha$ -PW<sub>11</sub>O<sub>39</sub>Zr( $\mu$ -OH)(H<sub>2</sub>O)}<sub>2</sub>] $\cdot$ 7H<sub>2</sub>O, **1**, was synthesized according to the procedure published in the literature.<sup>10b, 24</sup> D<sub>2</sub>O, DCl, NaOD, H<sub>3</sub>PW<sub>12</sub>O<sub>40</sub> $\cdot$ 23H<sub>2</sub>O, and the dipeptides used in this study were purchased from Sigma-Aldrich.

### NMR Measurements

<sup>1</sup>H NMR spectra were recorded on a Bruker Advance 400 spectrometer and TMSP-d<sub>4</sub> was used as an internal reference.

### Hydrolysis studies

The hydrolysis reaction mixtures typically contain 2.0 mM dipeptide, 2.0 mM **1**, and 0.5 mM TMSP-2,2,3,3-d<sub>4</sub> (sodium-3-trimethylsilylpropionate) in D<sub>2</sub>O. The pD of the reaction mixtures was adjusted to 5.4 with minor amounts of 1.0 M DCl or 1.0 M NaOD. The pH-meter reading was corrected by the equation: pD = pH + 0.41.<sup>25</sup> The reaction mixtures were kept at 60 °C and <sup>1</sup>H NMR spectra were taken after mixing and after different time increments. The hydrolysis products were identified by comparison of the <sup>1</sup>H NMR chemical shifts with those of the pure amino acids at the same pD value. The rate constants (*k*<sub>obs</sub>) were calculated by fitting the decrease in dipeptide concentration to a first-order mono-exponential decay function.

### Computational methods

*Binding study.* DFT calculations were performed with the ORCA program package,<sup>26</sup> version 3.0. Molecular geometries were fully optimized (without symmetry constraints) employing the gradient-corrected BP86<sup>27</sup> functional and def2-SVP basis sets<sup>28</sup>. MWB28 and MWB60 Stuttgart-Dresden effective core potentials<sup>29</sup> (ECPs) were used to replace the core electrons of Zr and W atoms respectively. To speed-up the calculations the Split-RI-J variant<sup>30</sup> of the resolution of the identity (RI) technique<sup>31</sup> with auxiliary def2-SVP/J Coulomb fitting basis sets<sup>32</sup> was employed (applying increased integration grid, Grid4 in the ORCA convention, and tight SCF convergence criteria). The effects of the aqueous solvent were included in the geometry optimizations by using the conductor-like screening model (COSMO),<sup>33</sup> as implemented in ORCA. The BP86 functional

has already been used for geometry optimization of various POMs in a number of theoretical works.<sup>34</sup> A comparison of the structural parameters for  $[\{\alpha\text{-PW}_{11}\text{O}_{39}\text{Zr}(\mu\text{-OH})(\text{H}_2\text{O})\}_2]^{8-}$ , **1**, obtained in this study and those determined by X-ray analysis of  $(\text{Et}_2\text{NH}_2)_8[\{\alpha\text{-PW}_{11}\text{O}_{39}\text{Zr}(\mu\text{-OH})(\text{H}_2\text{O})\}_2]\cdot 7\text{H}_2\text{O}$ <sup>24b</sup> in general showed a good agreement. Selected theoretical vs. experimental structural parameters may be found in Table S1 of the Supporting Information (SI). To find the lowest energy structures a conformational search was performed for all species under consideration (except **1** for which the X-ray structure was used as the input geometry). To verify that the stationary points found are minima on the potential energy surface (PES) and to obtain the thermochemical data numerical vibrational frequencies were computed using the standard rigid rotor-harmonic oscillator approximation. In order to refine the energies, single-point (SP) calculations on the BP86 optimized structures were performed by means of the hybrid density functional B3LYP<sup>35</sup> and more extended basis sets, def2-TZVP,<sup>28</sup> combined with the COSMO model. For the Zr and W atoms the above mentioned ECPs were still used. The Split-RI-J combined with the chain-of-spheres approximation to the exact exchange (RIJCOSX)<sup>36</sup> was utilized with the auxiliary def2-TZVP/J Coulomb fitting basis sets<sup>32</sup> (applying Grid5, GridX5 and tight SCF convergence criteria). When applied with the B3LYP functional the RIJCOSX approximation is expected to introduce a negligible error in the reaction energies as compared to the approximation free calculations.<sup>36</sup> To account for dispersion forces, missed in the pure B3LYP method, Grimme's atom-pairwise dispersion correction with Becke-Johnson damping (D3BJ)<sup>37</sup> was employed.

A detailed description of the strategy used for estimation of the free energies of complexation is given in section "Computational details" of the SI.

*Reaction path modeling.* This part of the calculations was performed with the Gaussian 09 Rev. D.01 suite of programs.<sup>38</sup> Geometry optimizations and frequency calculations were performed with the hybrid meta-GGA functional M06-2X<sup>39</sup> and 6-31+G(d,p) basis sets, as for the Zr and W atoms LANL2DZ<sup>40</sup> ECPs/basis sets were applied. Due to the large size of the transition state (TS) complexes intrinsic reaction coordinate (IRC) calculation were performed only in a few cases where we had doubts about the nature of the TS obtained. Otherwise, the TS structures were judged by analyzing the atomic movements of the corresponding imaginary frequencies. The bulk solvent effects were included in the geometry optimizations by means of the CPCM,<sup>41</sup> solvation model

with the default parameters (UFF cavities). However, the final free energies of solvation were obtained with Pauling atomic radii, taking into account the non-electrostatic energy contributions. The electronic energies of the stationary points were adjusted in SP calculations utilizing the M06-2X-D3, M06-D3,<sup>39</sup> and BMK-D3,<sup>42</sup> functionals (D3 means that the original D3 damping function<sup>37b</sup> is used instead of the BJ one) with def2-TZVP basis sets and MWB28/MWB60 ECPs for the Zr/W atoms. These functionals were selected on the basis of our previous study on POM catalyzed peptide bond hydrolysis.<sup>23</sup>

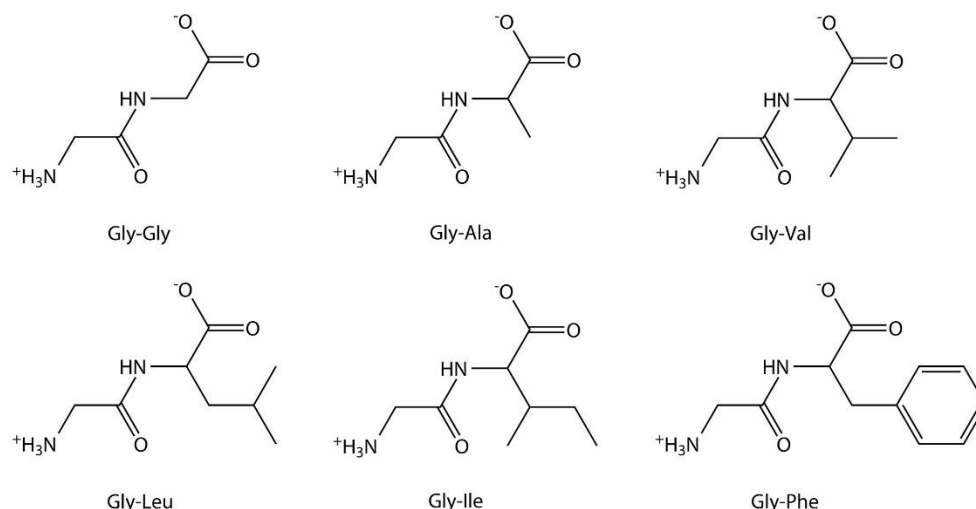
A standard state of 1 mol L<sup>-1</sup> was used for all compounds except water molecules for which 55.34 mol L<sup>-1</sup> standard state was used instead.<sup>43</sup>

## RESULT AND DISCUSSION

### Hydrolysis of Gly-X peptides

The hydrolysis of a series of Gly-X peptides in which X = Gly, Ala, Val, Leu, Ile and Phe (Scheme 1) in the presence of **1** was examined in solution containing equimolar amounts (2.0 mM) of the dipeptide and **1** at pD 5.4 and 60 °C. <sup>1</sup>H NMR spectra of the hydrolysis reactions were recorded at different reaction times and the hydrolysis product glycine (Gly) was identified by a singlet peak at 3.57 ppm<sup>10b</sup> (Figures S1 – S6 of the SI). The amount of the reacted dipeptide and of the free Gly was determined by integration of their respective <sup>1</sup>H NMR resonances and the rate constants (*k*<sub>obs</sub>) were calculated by fitting the decrease in dipeptide concentration to a mono-exponential decay function. The hydrolysis of all studied dipeptides in the absence of **1** under the same experimental conditions was also studied as control reaction. Fully homogeneous solutions were observed during the hydrolytic reaction of all dipeptides indicating that the presence of dipeptides does not affect the stability of **1**.<sup>10b, 19c</sup>





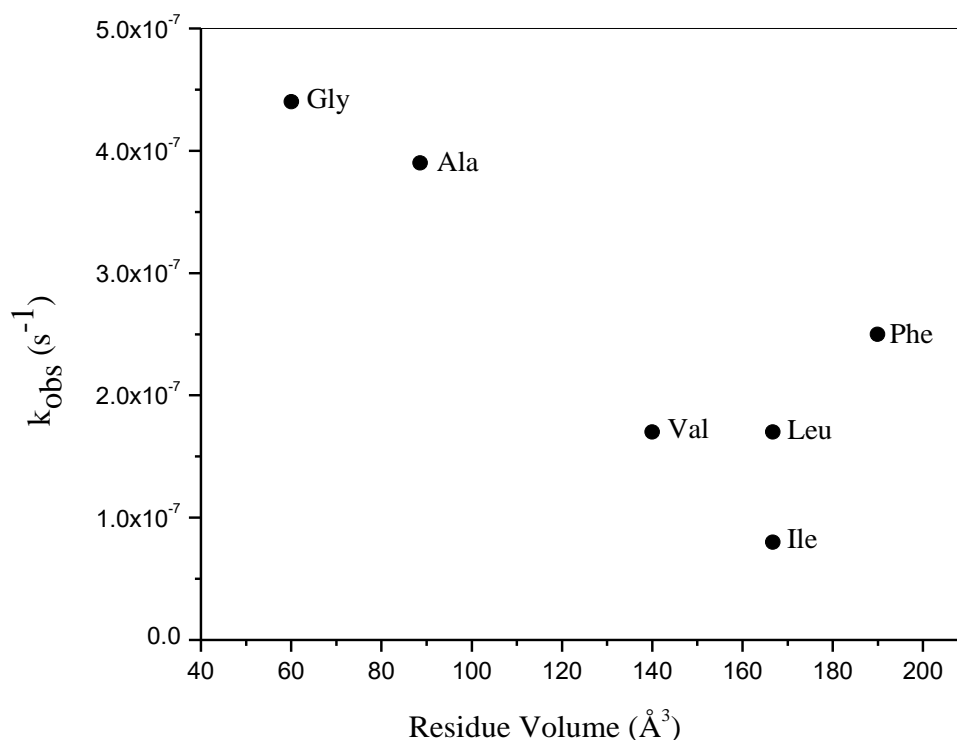
**Scheme 1.** Structures of Gly-X dipeptides.

The values of calculated rate constants for Gly-X hydrolysis are collected in Table 1. Due to the extreme stability of the peptide bond in the absence of **1**, the rate constants for the uncatalyzed hydrolysis of dipeptides could not be determined at these conditions. However, the amount of dipeptide hydrolyzed in the absence of **1** after seven months could be estimated and is also shown in Table 1. From these results it is clear that all dipeptides are very stable and that their hydrolysis was significantly accelerated by **1**. The mechanism of Gly-Gly hydrolysis in the presence of Zr-POMs was previously suggested to occur via coordination of Gly-Gly to a Zr(IV) ion via its amide oxygen and amine nitrogen atoms.<sup>10b, 23</sup> This interaction results in polarization of the amide C=O bond, making it more susceptible towards nucleophilic attack by water molecule.

**Table 1.** Observed rate constants for the hydrolysis of Gly-X in the absence and presence of **1** ([**1**] = [dipeptide] = 2.0 mM, pD 5.4, 60 °C).

Peptide	Without <b>1</b>	With <b>1</b>
	Conversion after 7 months (%)	$10^7 \times k_{\text{obs}} \text{ (s}^{-1}\text{)}$
Gly-Gly	25	$4.44 \pm 0.10$
Gly-Ala	5	$3.89 \pm 0.37$
Gly-Val	0	$1.71 \pm 0.19$
Gly-Leu	0	$1.72 \pm 0.35$
Gly-Ile	0	$0.81 \pm 0.11$
Gly-Phe	0	$2.53 \pm 0.15$

The dependence of the hydrolysis rate constants on the volume of the amino acid side chain<sup>44</sup> at the C-terminus is shown in Figure 2. From this figure it can be seen that the hydrolysis rate constant decreases when the volume of the non-coordinating X amino acid side chain is increased, implying that steric hindrance either impairs the interaction between the Zr(IV) ion and the dipeptide or prevents efficient nucleophilic attack of water or hydroxide. Interestingly, despite its higher molecular volume compared to Val, Ile and Leu, a higher rate constant was obtained for Gly-Phe as compared to Gly-Leu.



**Figure 2.** Reaction rate constants for the hydrolysis of 2.0 mM Gly-X dipeptides in the presence of 2.0 mM **1** at pD 5.4 and 60 °C as a function of the volume of amino acid X.

To shed light on the hydrolytic processes at the atomistic level we have first studied the coordination properties of the Gly-X dipeptides to the Zr(IV) atom(s) of the POM in order to discern the potential hydrolytically active species in solution. In a second step, the most likely candidates were used to model possible reaction paths of hydrolysis. Since in the series Gly-X the substrates differ only in their aliphatic (non-coordinating) residues we can safely assume that their

coordination properties will be rather similar. Thus, for the purpose of the binding study, we have performed a DFT investigation on the coordination behavior of the simplest dipeptide of this series, Gly-Gly.

## **Coordination properties of Gly-Gly towards Zr(IV)-substituted Keggin POMs**

### *Coordination of Gly-Gly to 2:2 Zr(IV)-Keggin POM (1)*

There are four different basic sites in the Gly-Gly zwitterion suitable for coordination to Zr(IV): the terminal amino nitrogen, the terminal carboxylic oxygens, the amide oxygen and the amide nitrogen. As previously demonstrated for Cu(I), Ni(I) and Co(I) ions, tri- and bidentate modes of Gly-Gly coordination are much more plausible than monodentate coordination modes. Also, the most plausible binding mode varies depending on the metal ion and is either bidentate (for Cu(I)), through the terminal amino and carboxylic groups or tridentate (for Ni(I) and Co(I)), involving in addition the amide nitrogen or oxygen atom.<sup>45</sup> It should be noted that in case of **1** tridentate binding to one metal center would result in a 9-coordinated zirconium, which is unlikely as the coordination numbers typical of Zr(IV) complexes including POMs are 7 or 8.<sup>46</sup> Due to steric hindrance caused by the bulky Keggin POM ligands our attempts to model even bidentate coordination of Gly-Gly to one Zr-atom of **1** were unsuccessful. Therefore, the most feasible bidentate modes of coordination were constructed, where the dipeptide bridges the two neighboring Zr-atoms, by varying the basic sites of Gly-Gly interacting with the metals. The optimized molecular structures, the free energies of complexation and their discussion can be found in Figure S7, Table S2 and section “Coordination of Gly-Gly to compound **1**” of the SI, respectively. In all cases positive free energies of complexation, in the range of 0.7 – 18.5 kcal mol<sup>-1</sup>, were computed. Thus, our DFT calculations predict that the formation of a complex between Gly-Gly and **1** is thermodynamically unfavorable. Only in the case of bidentate bridging coordination through the terminal carboxylic group a free energy of complexation close to zero was obtained (0.7 kcal mol<sup>-1</sup>), indicating that such a complex may occur in the reaction mixture. This result, however, cannot explain the hydrolytic activity of **1**. Moreover, the <sup>13</sup>C NMR data showed no evidence for -COO<sup>-</sup> binding, but instead suggest metal-ligand interaction involving the terminal amino and peptide carbonyl

groups.<sup>10b</sup> According to the calculations, however, formation of such a complex with **1** is rather unlikely, as this requires  $\sim 12$  kcal mol<sup>-1</sup> (at near neutral pH).

These results have stimulated us to find another explanation for the hydrolytic activity of **1**.

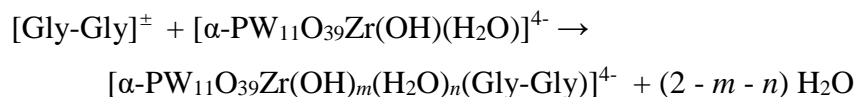
### *Coordination of Gly-Gly to 1:1 Zr(IV)-Keggin POM (2)*

Depending on the pH, temperature, time, concentration, and the presence of ligands the parent compound **1** may undergo partial dissociation in aqueous solution leading to the formation of both 1:2 species  $[\text{Zr}(\alpha\text{-PW}_{11}\text{O}_{39})_2]^{10-}$  with lacunary anion  $[\alpha\text{-PW}_{11}\text{O}_{39}]^{7-}$  and of monomeric species  $[\alpha\text{-PW}_{11}\text{O}_{39}\text{Zr}(\text{OH})_n(\text{H}_2\text{O})_{3-n}]^{(3+n)-}$  ( $n = 0$  and  $1$ ).<sup>11c, 26b</sup> The 1:2 complex appears in basic solution and shows a very low reactivity towards peptide bond hydrolysis,<sup>10b</sup> most probably because of the lack of free coordination sites at the Zr(IV) center when sandwiched between the two Keggin moieties. Therefore, substrate coordination to the 1:2 complex was not examined in our DFT models. The presence of monomeric species has first been suggested on the basis of <sup>31</sup>P NMR signals in strong acidic conditions (0.1 M aqueous HCl) by Nomiya et al.<sup>24b</sup> and in MeCN solution by Kholdeeva et al.<sup>11c</sup> Recently, Jiménez-Lozano et al. have calculated the energy of the reaction:  $2 [\text{PW}_{11}\text{O}_{39}\text{Zr}(\text{OH})(\text{H}_2\text{O})]^{4-} \rightarrow [\{\alpha\text{-PW}_{11}\text{O}_{39}\text{Zr}(\mu\text{-OH})(\text{H}_2\text{O})\}_2]^{8-}$ . Making use of PCM-B3LYP (with double- $\zeta$  basis sets and LANL2DZ ECPs for W and Zr) a dimerization energy of -23.6 kcal mol<sup>-1</sup> was obtained.<sup>47</sup> However, in terms of free energy in aqueous phase,  $\Delta G_{\text{aq}}^*$ , this dimerization energy is reduced to -3.4 kcal mol<sup>-1</sup>, suggesting that the monomeric species co-exists in equilibrium with **1**.<sup>22</sup> Finally, the presence of the monomer in solution at pH 6.4 was confirmed by means of <sup>31</sup>P NMR and <sup>31</sup>P DOSY experiments.<sup>22</sup> Even when present in minor concentrations the monomer could play a major role in the hydrolytic process. Indeed, formation of a stable monomer-substrate complex may shift the equilibrium in favor of the monomer.<sup>48</sup>

To further investigate this the coordination behavior and binding energy of Gly-Gly at the monomeric species were investigated more closely. First, the most stable monomeric form with respect to the coordination number and protonation state of H<sub>2</sub>O molecules to Zr was identified. In order to judge whether a Zr-OH<sub>2</sub> is deprotonated or not at the environmental pH the pK<sub>a</sub> value of a H<sub>2</sub>O ligand was estimated (by applying Eq. (9), see section “Computational details” of the SI) for the structure **2+H**, shown in Figure 3. It should be noted that the accurate estimation of pK<sub>a</sub>

values of highly charged species, such as POMs, is a challenging task by itself, implying more complicated thermodynamic cycles and approaches.<sup>49</sup> Nevertheless, the approach used here seems sufficiently reliable for the purpose of this study, as the calculated  $pK_a = 0.6$  is in good agreement with reported data for a Zr(IV) bound water molecule ( $pK_a \leq 0.6$ ).<sup>50</sup> Such a  $pK_a$  value enables the deprotonated form **2** to readily exist in neutral media. Coordination of an additional water molecule, giving rise to structures **2+Hw** and **2+w** in Figure 3, is found to be slightly endergonic both for **2+H** and **2**, by 1.4 and 1.7 kcal mol<sup>-1</sup> respectively. Therefore, structure **2** was used to obtain the free energy change upon coordination of Gly-Gly<sup>±</sup> to the monomeric Zr-POM.

As was done for structure **1**, many possible Gly-Gly complexes of **2** in their different protonation states were modeled. Only the lowest energy conformations are depicted in Figure 3. Note that, with exception of structure **2b0-H** (where the amide N is deprotonated and acts as the third ligand) either mono- or bidentate Gly-Gly coordination is obtained for all structures. The free energies of complexation ( $\Delta G_{aq}^{compl}$ ) were calculated according to the reaction

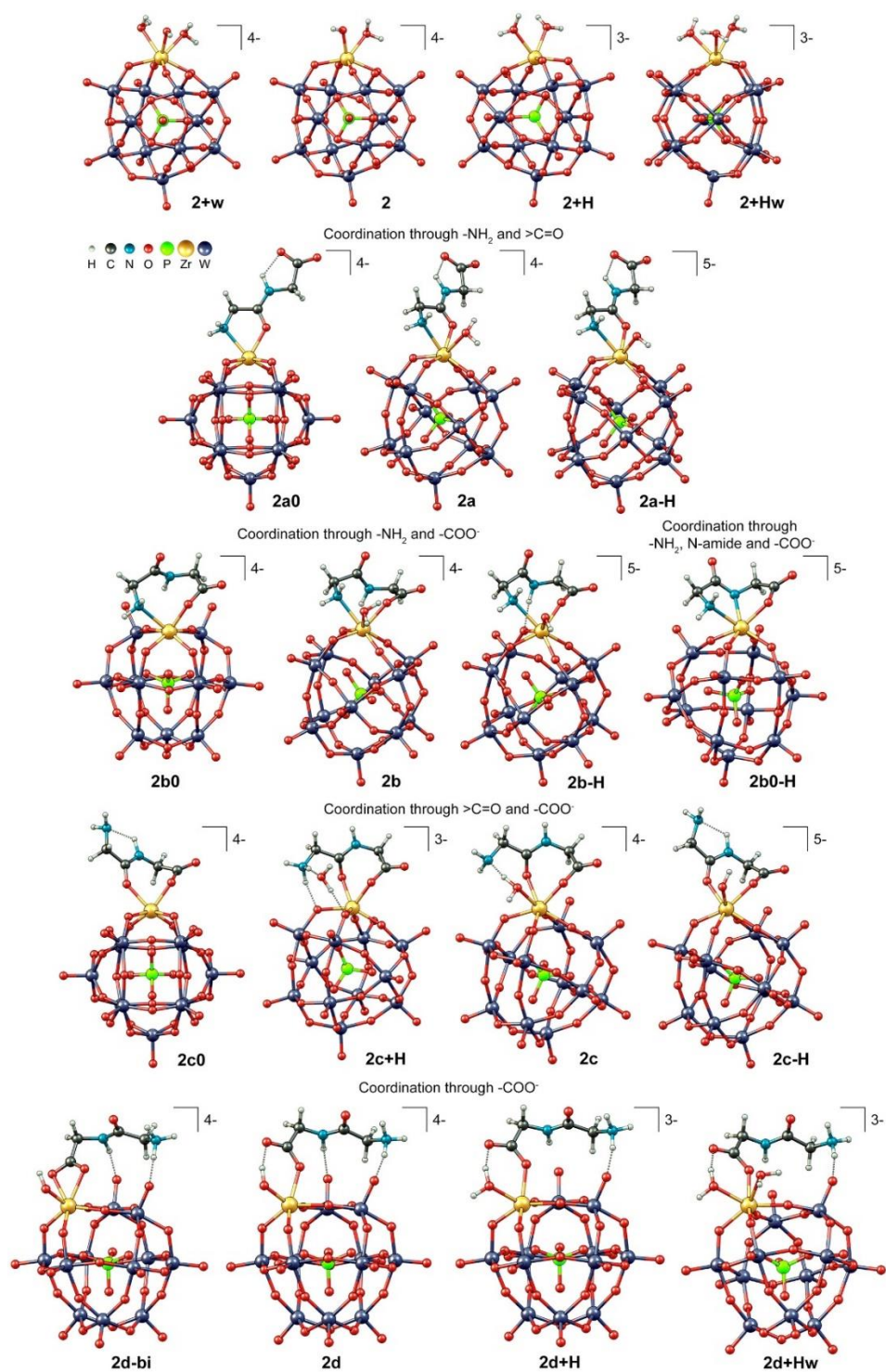


where, depending on the ligand binding mode  $m$  can be either 0 (where the Zr-OH ligand has received a proton from the substrate  $-\text{NH}_3^+$  group) or 1. For  $m = 0$ ,  $n = 0, 1$  or  $2$  while for  $m = 1$ ,  $n$  is either 0 or 1. Energy data are summarized in Table 2. The species shown in the left column of Table 2 are formed through the reaction above whereas the structures at the right hand side are obtained after deprotonation/protonation of the original reaction products. The formation free energies of the latter were obtained by adjusting the corresponding  $\Delta G_{aq}^{compl}$  values with the free energy change due to deprotonation/protonation at the environmental pH,  $\Delta G_{aq}^{compl}(\text{pH}) = \Delta G_{aq}^{compl} + \Delta G^{\text{deprot/prot}}(\text{pH})$ . More details can be found in section “Computational details” of the SI. The pH value of 6.4 was chosen for the binding studies to be able to compare the theoretical results with the suggested coordination mode of Gly-Gly to the POM based on <sup>13</sup>C NMR measurements (previously performed by us) at this pH.<sup>10b</sup> Coordination in the species labeled **a** occurs through the terminal amino and peptide carbonyl groups, in species **b** through the terminal amino and carboxyl groups, and in species **c** through the carbonyl and the carboxyl group. In the three species **2a**, **2b**, **2c**, Gly-Gly is coordinated as an anion. These structures are obtained from **2**

by protonation of Zr-OH by the ligand  $\text{-NH}_3^+$  group and simultaneous ligand exchange (water release). In structures **2a0**, **2b0**, **2c0** the remaining  $\text{H}_2\text{O}$  is also dissociated. As can be seen from Table 2, all three types of Gly-Gly coordination are found to be exergonic, releasing an energy between 0.9 and 5.2 kcal mol<sup>-1</sup>. In the species labeled **d**, Gly-Gly coordination involves only the substrate  $\text{-COO}^-$  group, either mono- (**2d**) or bidentate (**2d-bi**). Here, Gly-Gly is coordinated as a zwitterion, and has replaced a water ligand in structure **2**. Bidentate coordination is only slightly more stable than monodentate coordination, by 0.6 kcal mol<sup>-1</sup>, and both structures are similar in energy to the alternative coordination modes **a**, **b**, **c**.

As all three species **2a**, **2b**, **2c** contain a  $\text{H}_2\text{O}$  ligand, deprotonation of this ligand might be favorable, giving rise to the species **2a-H**, **2b-H**, **2c-H**. The corresponding  $\text{pK}_a$  value were found to be 3.8, 10.0, and 11.1. This means that further stabilization by deprotonation is only expected for species **2a**, with a calculated energy gain (at pH 6.4) of -3.6 kcal mol<sup>-1</sup>, giving rise to a net formation energy ( $\Delta G_{\text{aq}}^{\text{compl}}(\text{pH})$ ) of -4.5 kcal mol<sup>-1</sup> for complex **2a-H**. On the other hand, species **2d** and **2d-bi** contain an -OH ligand. The basicity of this -OH was also estimated. Starting from the most stable **2d-bi** structure, a  $\text{pK}_b$  value of 3.7 was calculated, corresponding to a gain in free energy of 5.4 kcal mol<sup>-1</sup> for protonation (at pH 6.4). Note that protonation gives rise to a complex with a mono-coordinated carboxylate group, **2d+H**. The corresponding bidentate structure was found to be less stable by 5 kcal mol<sup>-1</sup> and is not included in Figure 3. Coordination of a second water molecule to **2d+H** does not lead to additional stabilization of this complex (structure **2d+Hw**).

Finally, two other acid-base reactions were found to give rise to structures with similar energies as those already considered. The first is structure **2b0-H**, obtained from deprotonation of the amide nitrogen in **2b0**. As Figure 3 shows, **2b0-H** is a tridentate complex. However, its formation from **2b0** is endergonic by 6.7 kcal mol<sup>-1</sup> ( $\text{pK}_a = 11.3$ ). The second structure is **2c+H**, obtained from **2c** by protonation of the endstanding ligand  $\text{-NH}_2$  group. With a  $\text{pK}_b$  of 7.0, protonation is slightly favored, by 0.8 kcal mol<sup>-1</sup>, at pH 6.4. However, one can observe that in structure **2c+H** the Zr-OH<sub>2</sub> bond is broken, and a HO-H...O-Zr H-bond is formed instead.



**Figure 3.** DFT optimized structures of the Gly-Gly complexes with **2**.

In order to compare the results obtained for **2** with those for the initial compound **1** (Table S2) the free energy change corresponding to the net reaction (Dimer) + 2 GlyGly  $\rightarrow$  2 (Monomer complex) should be used. These values are provided in parentheses in Table 2. Even though these formation energies must be considered more or less qualitatively (e.g. an error of 1 pK<sub>a</sub> value introduces an error of 1.4 kcal mol<sup>-1</sup> in the  $\Delta G_{\text{aq}}^{\text{compl}}$ (pH)), our calculations undoubtedly show that substrate coordination to the monomeric POM is largely preferred as compared to the initial dimer.

**Table 2.** Formation free energy (kcal mol<sup>-1</sup>) of the coordination compounds of Gly-Gly and **2**.

Species <sup>a</sup>	$\Delta G_{\text{aq}}^{\text{compl}}$	pK <sub>a</sub> /pK <sub>b</sub>	$\Delta G^{\text{deprot/prot}}$	Species <sup>b</sup>	$\Delta G_{\text{aq}}^{\text{compl}}$ (pH)
<b>2a0</b>	-1.1 (1.2) <sup>c</sup>				
<b>2a</b>	-0.9 (1.6)	<b>3.8</b>	-3.6	$\rightarrow$ <b>2a-H</b>	-4.5 (-5.6)
<b>2b0</b>	-5.2 (-7.0)	<b>11.3</b> (>NH)	6.7 (>NH)	$\rightarrow$ <b>2b0-H</b>	1.6 (0.2)
<b>2b</b>	-4.2 (-5.0)	<b>10.0</b>	5.0	$\rightarrow$ <b>2b-H</b>	0.8 (1.8)
<b>2c0</b>	-2.5 (-1.6)				
<b>2c</b>	-1.9 (-0.4)	7.0 (NH <sub>2</sub> )	-0.8 (-NH <sub>2</sub> )	$\rightarrow$ <b>2c+H</b>	-2.8 (-2.2)
		<b>11.1</b>	6.4	$\rightarrow$ <b>2c-H</b>	4.4 (12.2)
<b>2d</b>	-1.9 (-0.4)				
<b>2d-bi</b>	-2.5 (-1.6)	3.7	-5.4	$\rightarrow$ <b>2d+H</b>	-7.9 (-12.4)
				<b>2d+Hw</b>	-6.8 (-10.2)

<sup>a</sup> Coordination compounds formed according to the reaction given in the text (for more details see section “Computational details” of the SI).

<sup>b</sup> Species obtained via deprotonation/protonation of the corresponding compounds in the first column (for more details see section “Computational details” of the SI).

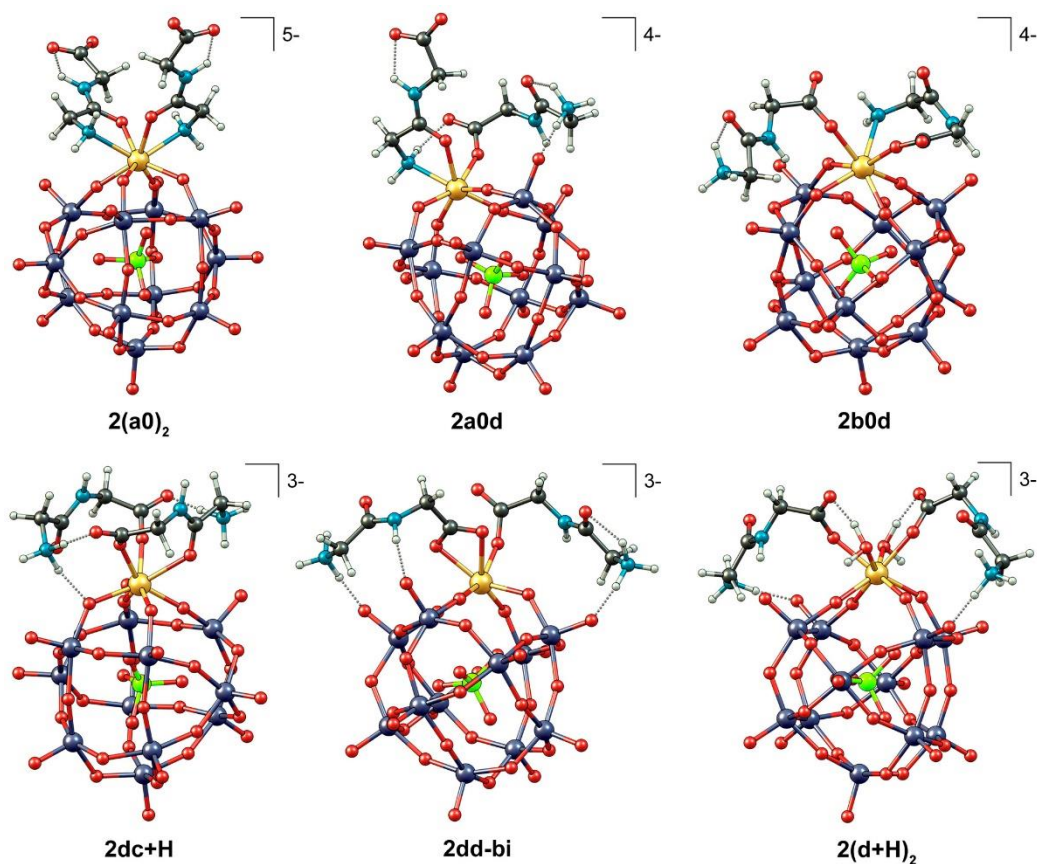
<sup>c</sup> Energy values in parentheses are relative to the parent compound **1**, assuming that both generated monomeric species **2** react with Gly-Gly ligands.



### *1:1 Zr(IV)-Keggin POM complexes with two Gly-Gly ligands*

In case of **2** coordination of a second Gly-Gly ligand to the Zr(IV) center is also possible. To examine this scenario a series of complexes were modeled by testing different binding modes of Gly-Gly to the **2a0**, **2b0**, **2c+H** and **2d+H** species. These species were selected based on their stability, for every particular binding mode, and the availability of free coordination sides at the Zr(IV) center (six-coordinated zirconium ion). The resultant optimized structures and their formation energies are given in Figure 4 and Table 3 respectively. With exception of **2b0d** and **2(d+H)<sub>2</sub>** the addition of a second Gly-Gly ligand to **2a0**, **2b0**, **2c+H** and **2d+H** structures (with or without H<sub>2</sub>O release) was found as exergonic by 0.4 (**2dd-bi**) – 8.0 (**2a0d**) kcal mol<sup>-1</sup>. However, comparison of the formation free energies in Table 2 and Table 3 reveals that there is no substantial stabilization effect upon coordination of a second substrate molecule to **2** instead of a H<sub>2</sub>O/OH ligand. Thus, the calculations predicted that in the reaction mixture Gly-Gly complexes with **2** involving either one or two substrate molecules may co-exist in equilibrium. Among all calculated complexes the lowest energy structure, **2a0d** ( $\Delta G_{\text{aq}}^{\text{compl}} = -9.1$  kcal mol<sup>-1</sup>), contains two Gly-Gly molecules, one coordinated through the –C=O and –NH<sub>2</sub> groups and the other through the –COO<sup>-</sup> group (Figure 5). Energy contributions to the aqueous free energy of all complex species can be found in Table S3.

Based on the results thus obtained, the **2a0d** and the **2a-H** species may be considered as the most probable candidates responsible for the hydrolytic activity of **1**. In both complexes the peptide bond of the bidentate bound Gly-Gly ligand is activated towards nucleophilic attack through polarization of the –C=O bond by the Zr(IV) ion. On the one hand **2a0d** reveals the highest stability while on the other hand **2a-H** contains an intramolecular -OH nucleophile which is a good premise for high hydrolytic ability of this species. These two complexes were used for the mechanistic investigations in the next section.



**Figure 4.** DFT optimized structures with two Gly-Gly ligands coordinated to **2**.

**Table 3.** Formation free energy ( $\text{kcal mol}^{-1}$ ) of the complexes obtained through coordination of two Gly-Gly ligands to **2**.

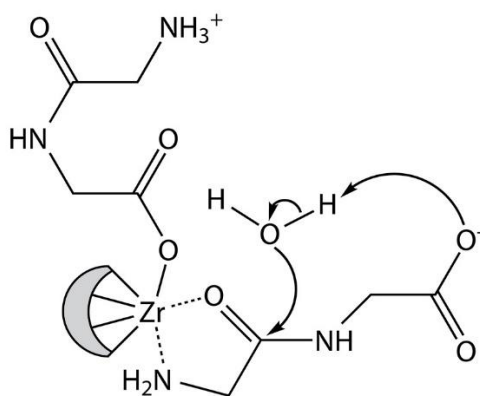
Species	$\Delta G_{\text{aq}}^{\text{compl}}/\Delta G_{\text{aq}}^{\text{compl}}(\text{pH})$
<b>2(a0)<sub>2</sub></b>	-3.6 (-3.9) <sup>a</sup>
<b>2a0d</b>	-9.1 (-14.8)
<b>2b0d</b>	-1.3 (0.7)
<b>2dc+H</b>	-7.6 (-11.8)
<b>2dd-bi</b>	-8.2 (-13.2)
<b>2(d+H)<sub>2</sub></b>	0.7 (4.8)

<sup>a</sup> Energy values in parentheses are relative to the parent compound **1**, assuming that both generated monomeric species **2** react with two Gly-Gly ligands.

## Mechanism of Gly-Gly hydrolysis catalyzed by the monomeric compound **2**.

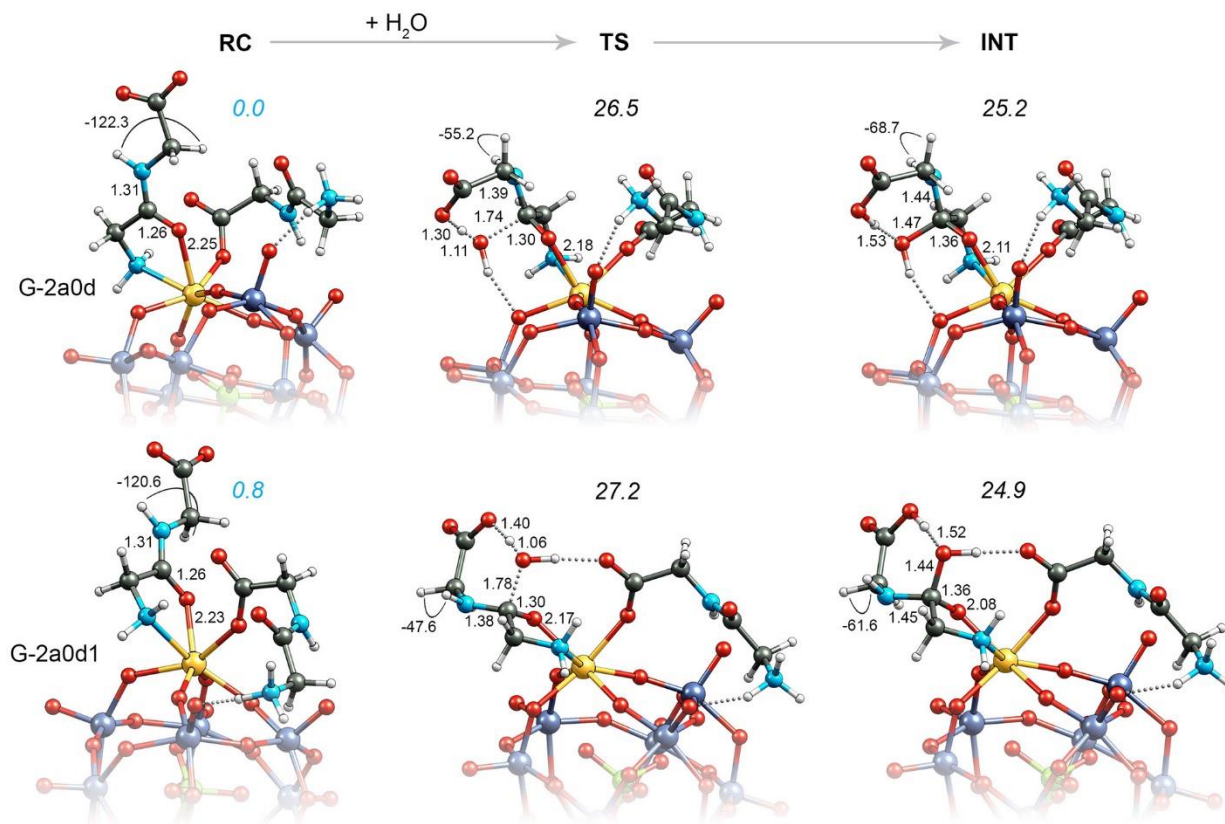
Free energy surfaces of the possible mechanistic scenarios were probed by means of M06-2X, M06 and BMK density functionals. The results obtained with the three functionals are in good qualitative agreement. Therefore, unless otherwise noted, only the M06 energy values will be discussed below.

In a recent theoretical investigation on the mechanism of Gly-Gly hydrolysis catalyzed by a dimeric tetrazirconium(IV) Wells-Dowson POM<sup>23</sup> we have shown that in the hydrolytically active complex the substrate is bidentate coordinated to one Zr(IV) center, by analogy with **2a0d**, and that nucleophilic attack by a water molecule on the amide C-atom is assisted by the ligand -COO<sup>-</sup> group. This water addition step was found to be the bottleneck of the whole process of hydrolysis.<sup>23</sup> A similar mechanism could be proposed for the **2a0d** species while in the **2a-H** species the intramolecular Zr-OH attack dominates over the -COO<sup>-</sup> assisted hydrolysis. There are two ligand -COO<sup>-</sup> groups in the **2a0d** complex which potentially could act as a general base, one is Zr-bound while the other is free. Since the first is a weaker proton acceptor than the latter the hydrolytic mechanism here most likely involves nucleophilic attack by a solvent water molecule with proton abstraction by the free -COO<sup>-</sup> group of Gly-Gly, as shown in Scheme 2. This reaction path will be further referred to as “2a0d path”.



**Scheme 2.** Principle mechanism of nucleophilic attack on the amide carbon atom of Gly-Gly in the **2a0d** complex.

Because water addition to the amide C-atom is expected to be the rate-determining step along the 2a0d path our efforts were mainly focused on modeling this reaction step. The **2a0d** species itself encompasses several conformations with similar stabilities corresponding to different mutual orientations of the two Gly-Gly ligands and the POM, and the preferential stability order of these conformations changes with the applied functional. Moreover, different reactant complexes (RC) are connected to different TSs for nucleophilic attack on the PES. To deal with this complex situation we have considered several initial conformations, and their associated TSs, instead of just the one RC conformation with the lowest energy, as predicted by a certain functional. The number of conformations was limited by selecting only those whose relative stabilities, calculated with the M06-2X functional, lie within 2 kcal mol<sup>-1</sup>. In this way the lowest energy **2a0d** conformations predicted by the three functionals are included and examined. Two RC conformations, **G-2a0d** (as also predicted by B3LYP in the binding study) and **G-2a0d1** were selected, and are depicted in Figure 5. In the lowest energy TS connected to **G-2a0d** on the PES the attacking water molecule is H-bonded to a bridging oxygen atom of the POM skeleton whilst the free ligand -COO<sup>-</sup> group abstracts a H<sub>2</sub>O proton and the nascent OH nucleophile attacks the amide carbon atom. This process require 26.5 kcal mol<sup>-1</sup>. The TS relaxes to an unstable intermediate INT with an energy lowering of only 1.3 kcal mol<sup>-1</sup> (Figure 5). On the other hand, in the TS structure connected to **G-2a0d1** the attacking water molecule is H-bonded to a Zr-O-C=O oxygen atom of the neighboring Gly-Gly ligand. The calculated energy barrier associated with this process is 27.2 kcal mol<sup>-1</sup>, and a similar TS → INT energy lowering of 2.3 kcal mol<sup>-1</sup> was also found. Hence, in average the 2a0d path of hydrolysis requires an activation energy of around 27 kcal mol<sup>-1</sup>. It should be noted that the three functionals used here tend to underestimate the INT stability<sup>23</sup> which may explain the small TS – INT energy difference. Additional energy data are provided in Table S4.



**Figure 5.** Optimized RC, TS and INT structures involved in the rate-determining step of Gly-Gly hydrolysis via the 2a0d path. Bond lengths, dihedral angles and relative free energies (in *Italics*) are in Å, degrees and kcal mol<sup>-1</sup>, respectively. The energy values colored in cyan indicate the relative stability of the RC complexes while the TS and INT energies (in black) are calculated with respect to the energies of the corresponding starting RC conformations and a free H<sub>2</sub>O molecule. All energy values are obtained with the M06 functional.

The hydrolytic mechanism originating from the **2a-H** species, further referred to as “2a-H path”, was also thoroughly examined. The results are presented and discussed in section “Gly-Gly hydrolysis via 2a-H path” of the SI (Figures S8 – S10 and Tables S5 and S6). As expected, intramolecular OH attack on the amide C-atom with formation of a tetrahedral intermediate proceeds with a low energy barrier, estimated at 18.8 kcal mol<sup>-1</sup>. The rate-determining step along this path, however, is the last reaction step corresponding to a water mediated proton transfer from the –OH group to the amide nitrogen atom with a subsequent substrate decomposition. This process requires 23.8 kcal mol<sup>-1</sup>, which is lower than the activation energy associated with the 2a0d path by about 3 kcal mol<sup>-1</sup>.

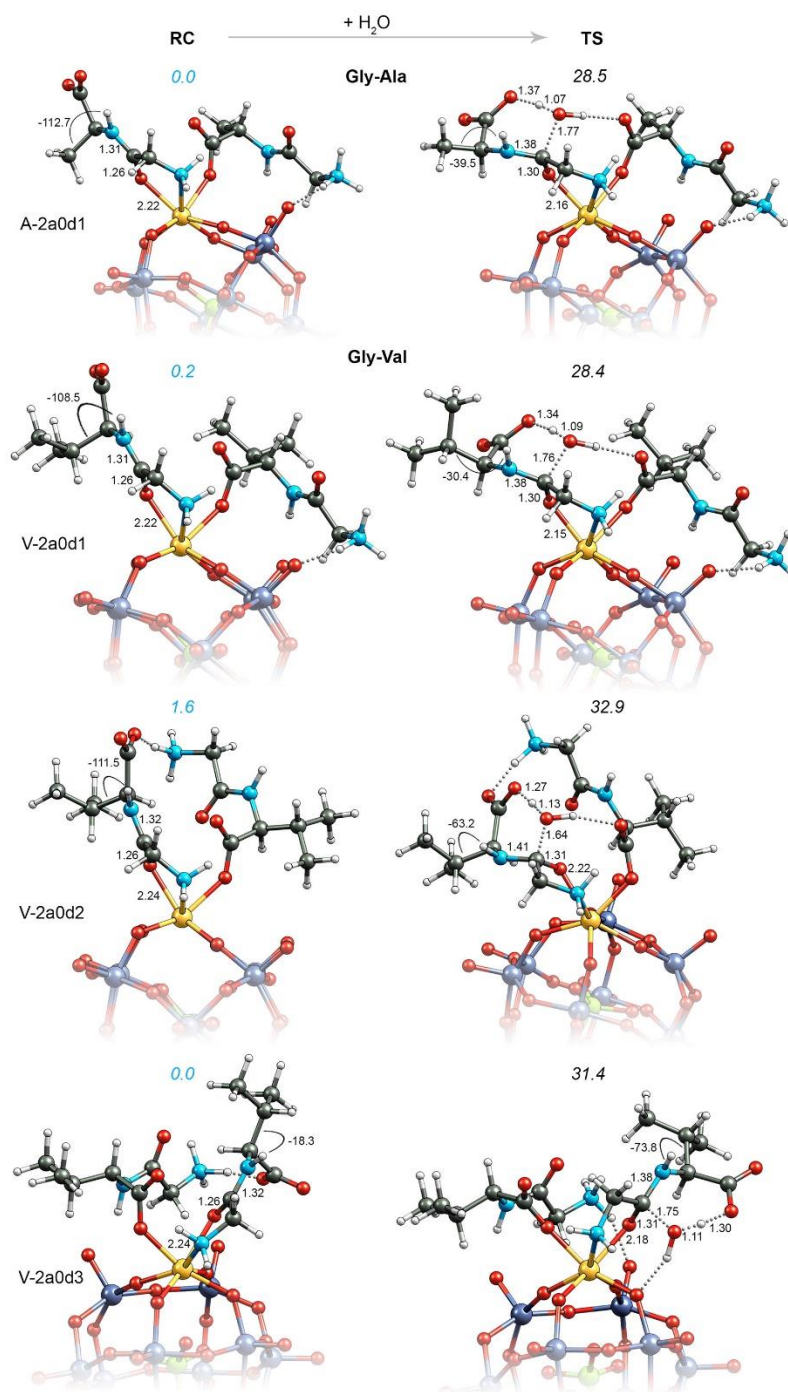
The experimental  $\Delta H^\ddagger$  and  $\Delta S^\ddagger$  values for Gly-Gly hydrolysis in the presence of **1** were previously estimated to be 20.9 kcal mol<sup>-1</sup> and -29.4 cal mol<sup>-1</sup> K<sup>-1</sup>, giving  $\Delta G^\ddagger(25\text{ }^\circ\text{C}) = 28.3\text{ kcal mol}^{-1}$ .<sup>10b</sup> The predicted activation energy for the 2a0d path (27 kcal mol<sup>-1</sup>) is in better agreement with the experiment than that for the 2a-H path, thus suggesting the 2a0d species/mechanism as responsible for the observed peptide hydrolysis. Interestingly, the rate of Gly-Gly hydrolysis in the presence of the dimeric tetrazirconium(IV) Wells-Dowson POM,  $k_{\text{obs}} = 2.97 \times 10^{-7}\text{ s}^{-1}$ , is very similar to the one determined here ( $4.44 \times 10^{-7}\text{ s}^{-1}$ ).<sup>23</sup> This POM, however, does not contain a Zr-OH function(s) and in contrast to **1** it is stable in a broad range of experimental conditions.<sup>23</sup> In addition, the changes in the rate constants within the Gly-X series of dipeptides observed with the two catalysts follow the same trend. These findings strongly suggest that the mechanisms of Gly-X peptide bond hydrolysis catalyzed by the two POMs are rather similar, pointing out that the 2a0d path (which is analogous to the one proposed for the Zr<sub>4</sub>-POM) is indeed the most likely mechanism of hydrolysis.

The activation energy difference between 2a0d and 2a-H paths (1-3 kcal mol<sup>-1</sup> with all three functionals, see Tables S4 and S6) implies that hydrolysis should occur faster by a few orders of magnitude along the 2a-H path. It has been shown that Zr(IV) complexes delivering an intramolecular OH nucleophile to the peptide bond are capable of hydrolyzing 90% of Gly-Gly within 20h, at neutral pH and 60 °C,<sup>51</sup> which is significantly faster than the reaction rates observed here (with half-life of more than 2 weeks). Taking into account the calculated difference in the stabilities of the two complexes, **2a-H** (-4.5 kcal mol<sup>-1</sup>, Table 2) and **2a0d** (-9.1 kcal mol<sup>-1</sup>, Table 3), and the possible existence of other species, more stable than **2a-H** (e.g. **2d+H**, **2dc+H** and **2dd-bi**), it seems evident that the equilibrium concentration of **2a-H** in the reaction mixture is probably too low to influence the hydrolysis rate.

### Effect of the $\alpha$ -substituent on the POM catalyzed hydrolysis of dipeptides with aliphatic side chains.

Due to the similarity of the Gly-X substrates and their rates of hydrolysis the findings obtained for Gly-Gly should also be valid for the rest of dipeptides in the series. To trace the effects of the aliphatic side chains on the reaction kinetics the hydrolysis of Gly-Ala and Gly-Val via the 2a0d

path was examined. Unfortunately, the comparatively higher side chain flexibility of Gly-Leu, Gly-Ile and Gly-Phe, in combination with the variety of different mutual orientations of the two ligands and the POM in the complexes, requires an extensive DFT conformational search which substantially increases the computational cost. Therefore substrate-catalyst complexes and TSs with these three ligands were not modeled. It should be mentioned that the differences between the rate constants in the Gly-X series are small, corresponding to activation free energy differences smaller than 1 kcal mol<sup>-1</sup>. This is below the accuracy of the present computational chemistry approaches. However, since Gly-Gly, Gly-Ala and Gly-Val differ only in the size of the aliphatic residue and the nature of both RCs and TSs is the same for the three ligands the predicted trends should be reliable due to error cancelation. By analogy with Gly-Gly, only RC conformations with relative stabilities less than 2 kcal mol<sup>-1</sup> were considered. For the Gly-Ala complex of **2** the three density functionals agree that **A-2a0d1** (Figure 6), which is analogous to **G-2a0d1**, is the most stable conformation. The corresponding TS is also analogous to the one located for Gly-Gly, but the free energy barrier now amounts to 28.5 kcal mol<sup>-1</sup>. In the case of Gly-Val three reactant complexes were selected: **V-2a0d1**, **V-2a0d2** and **V-2a0d3** (Figure 6). Complex **V-2a0d1** is analogous to **G-2a0d1** and **A-2a0d1** while in the complexes **V-2a0d2** and **V-2a0d3** a H-bond is formed between the -NH<sub>3</sub><sup>+</sup> and the -COO<sup>-</sup> groups of the two ligands. The calculated free energy barriers for nucleophilic addition of a water molecule in the **V-2a0d1**, **V-2a0d2** and **V-2a0d3** forms are 28.4, 32.9 and 31.4 kcal mol<sup>-1</sup>, respectively (or 31 kcal mol<sup>-1</sup> in average). Thus, in average the calculations predict an increase of the activation energy (decrease in the rate constants) in the order Gly-Gly < Gly-Ala < Gly-Val which is in full agreement with experiment. Activation energies calculated with the M06-2X and BMK functionals are collected in Table S7.

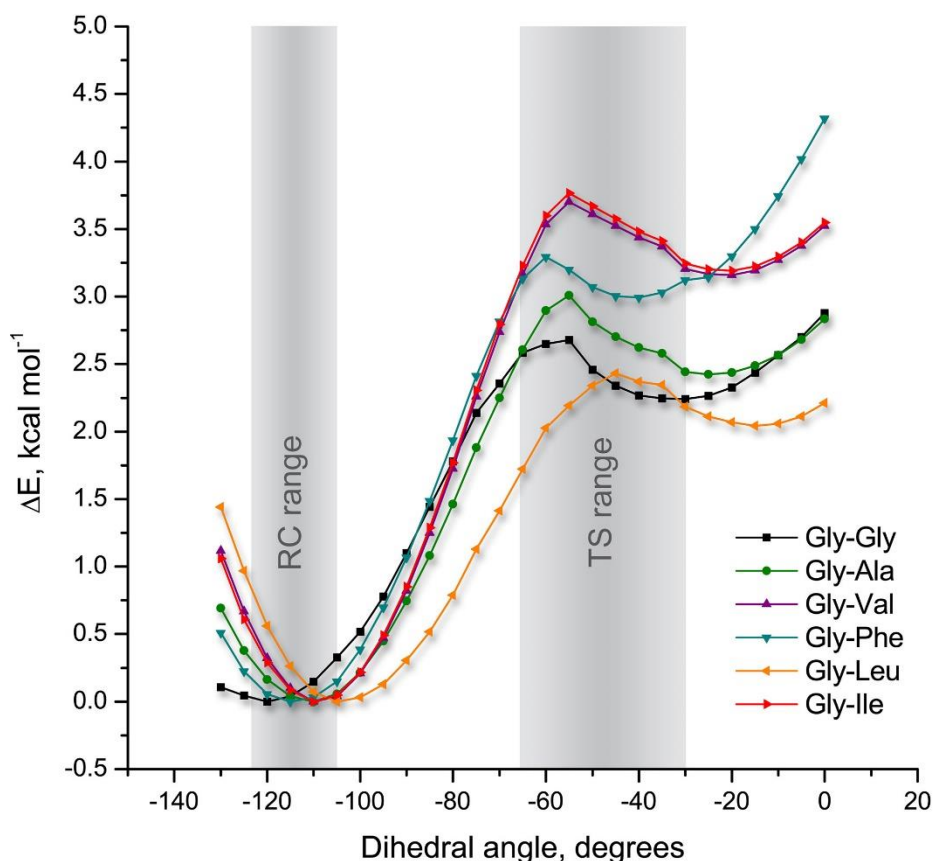


**Figure 6.** Optimized RC and TS structures involved in the rate-limiting hydrolytic step of Gly-Ala and Gly-Val when bound to **2** through **2a0d** binding mode. Bond lengths, dihedral angles and relative free energies (in *Italics*) are in Å, degrees and kcal mol<sup>-1</sup>, respectively. The energy values colored in cyan indicate the relative stability of the RC complexes while the TS energies (in black) are calculated with respect to the energies of the corresponding starting RC conformation and a free H<sub>2</sub>O molecule. All energy values are obtained with the M06 functional.



A closer look at the TS structures in Figure 5 and Figure 6 reveals that in all cases the aliphatic side chains do not hinder the attacking nucleophile. Indeed, the distance between the H<sub>2</sub>O oxygen and the closest C-atom of the side chain is always larger than 4.4 Å. Thus, the experimentally observed reaction rate decrease in the Gly-X series cannot be explained in terms of steric hindrance of the attacking nucleophile. Interestingly, we found that the decrease in the hydrolysis rate in presence and in absence of **1** follow the same trend. The uncatalyzed reactions are, however, too slow at these experimental conditions to determine the corresponding rate constants (Table 1). To describe the uncatalyzed hydrolysis in terms of activation energies the water addition step to the free Gly-Gly, Gly-Ala and Gly-Val dipeptides was also modeled (the mechanism of the uncatalyzed Gly-Gly hydrolysis was previously reported by us<sup>23</sup>). In line with experiment the calculated activation free energies of Gly-Gly, Gly-Ala and Gly-Val are 32.0, 32, 6 and 34.1 kcal mol<sup>-1</sup>, respectively. Relevant structures and energy data can be found in Figure S11 and Table S8. These results suggest that the trend in the observed rate constants of the POM catalyzed reactions might be related to an inherent property of the substrates. To check whether the  $\alpha$ -substituent can influence the electron density distribution of the peptide bond and potentially alter the susceptibility of the amide carbon towards nucleophilic attack a number of molecular descriptors, e.g. NBO analysis, Hirshfeld charges and atomic electrostatic potential, were computed for all dipeptides in the Gly-X series. The results showed insignificant changes in the peptide bond properties when changing the aliphatic side chain, see Table S9. A reasonable explanation, however, can be found on the basis of the substrate conformational changes in the complexes of Gly-Gly, Gly-Ala and Gly-Val upon activation. As can be seen in Figure 5 and Figure 6 the dihedral angle H-N-C $\alpha$ -H $\alpha$  (where H $\alpha$  is the  $\alpha$ -hydrogen atom being substituted in the L enantiomers of the corresponding dipeptides) decrease, in absolute values, from -122° ÷ -105° in the RCs to -60° ÷ -30° in the TS structures. Substitution of H $\alpha$  in Gly-Gly with a bulky aliphatic substituent is expected to increase the steric hindrance between the amide hydrogen and the  $\alpha$ -substituent in the rate-limiting transition state, thus increasing the activation energy for hydrolysis. To check this assumption the energy change of the free Gly-Gly, Gly-Ala, Gly-Val, Gly-Leu, Gly-Ile and Gly-Phe dipeptides as a function of the H-N-C $\alpha$ (or H $\alpha$ ) angle was examined by means of M06-2X functional. The results are plotted in Figure 7. As can be seen in Figure 7, within the range -60° ÷ -30° (TS range) the rotational barrier increases in the order Gly-Gly < Gly-Ala < Gly-Phe < Gly-Val < Gly-Ile, which is in perfect agreement with the experimentally observed decrease

in the rate constants of the POM catalyzed reactions. The only disagreement with the experimental trend was found for Gly-Leu with rotational energy barrier lower than that of Gly-Gly. Obviously, other factors such as weak intramolecular ligand-ligand and/or ligand-POM interactions, which could additionally stabilize/destabilize either the initial RC or the activated TS states, also contribute to the barrier height.



**Figure 7.** Energy profile of the H-N-C-C<sub>α</sub>(or H) dihedral angle rotation in Gly-Gly, Gly-Ala, Gly-Val, Gly-Leu, Gly-Ile and Gly-Phe dipeptides obtained with relaxed PES scan at the CPCM-M06-2X/6-311+G(2df,2p) level of theory. The ranges of calculated dihedral angles for both RCs and TSs of the Gly-Gly, Gly-Ala and Gly-Val complexes with **2** (**2a0d** binding mode) are indicated as “RC range” and “TS range”, respectively.

## CONCLUSIONS

A detailed examination of the effect of the bulkiness of amino acid side chain on the hydrolysis rate of a series of dipeptides with Gly-X sequence in the presence of a dimeric Zr(IV)-substituted Keggin POM was performed by means of a combined theoretical-experimental approach. The observed rate of peptide bond hydrolysis was found to decrease with increase of the side chain bulkiness in the order Gly-Gly > Gly-Ala > Gly-Phe > Gly-Leu  $\approx$  Gly-Val > Gly-Ile. Experimental data, reported previously by us,<sup>10b</sup> evidenced that a Gly-Gly-catalyst complex involving coordination of both amide  $\text{-C=O}$  and terminal  $\text{-NH}_2$  groups to the Zr(IV) center of the POM is formed prior hydrolysis. DFT calculations, presented here, revealed that coordination of the substrate to the starting compound **1** is thermodynamically not favored, and that Gly-Gly instead binds to the monomeric POM derivative **2** which is also present in equilibrium with **1**. The formation free energy associated with the experimentally suggested coordination mode of Gly-Gly to **1** is calculated to be rather endergonic,  $11.9 \text{ kcal mol}^{-1}$  (at pD 6.4). In contrast, bidentate coordination through the  $\text{-C=O}$  and  $\text{-NH}_2$  groups to **2** was predicted to be exergonic by  $-5.6 \text{ kcal mol}^{-1}$ . The stability of this species is further increased upon binding of a second Gly-Gly ligand, through the terminal  $\text{-COO}^-$  group, as the net complexation free energy becomes  $-14.8 \text{ kcal mol}^{-1}$ . The latter is considered as the major player in the peptide bond cleavage. The mechanism of hydrolysis involves nucleophilic attack by a solvent water molecule on the amide carbon atom of the bidentate bound ligand. This process is assisted by the free  $\text{-COO}^-$  group of the same ligand which acts as a general base to abstract a proton from the attacking water molecule. Since the uncatalyzed hydrolysis of Gly-X dipeptides is also assisted by the carboxylic group, the main catalytic role of the POM can be attributed to the electrophilic activation of the carbonyl moiety by the Zr(IV) ion. The alternative mechanism involving intramolecular nucleophilic attack by a Zr-OH group was ruled out because this mechanism implies hydrolysis rates significantly higher than that observed in this study.

The decrease of the hydrolysis rate with increase of the side chain bulkiness is mostly due to the increased ligand conformational strain in the rate-limiting transition state which elevate the reaction activation energy. This conformational strain is imposed by the steric hindrance between the amide hydrogen atom and the  $\alpha$ -substituent. It increases first, upon substitution of  $\text{H}_\alpha$  in Gly-Gly with aliphatic  $\alpha$ -substituent and second, with the  $\beta$ -branching of the  $\alpha$ -substituent. This explanation should also be valid for the hydrolysis of Gly-X in presence of dimeric

tetrazirconium(IV) Wells-Dowson POM where the same trend in the rate constants was observed. This work demonstrates the power of a combined theoretical and experimental study to unravel the molecular origin of the catalytic activity of metal-substituted POMs towards organic transformations, especially since reactivity studies on this type of compounds are often hampered by the rich solution chemistry of metal-substituted POMs.

## **ASSOCIATED CONTENT**

### **Supporting Information**

<sup>1</sup>H NMR spectra, additional computational details, a discussion of Gly-Gly binding to **1**, energy contributions to the aqueous free energies of the Gly-Gly-POM complexes, a discussion of 2a-H path of hydrolysis, energy contributions to the relative free energies of the stationary points along the reaction paths calculated with the M06, M06-2X and BMK functionals, molecular structures of relevant species in Cartesian coordinates. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## **AUTHOR INFORMATION**

### **Corresponding Author**

\*E-mail: [tzvetan.mihaylov@kuleuven.be](mailto:tzvetan.mihaylov@kuleuven.be)

## **ACKNOWLEDGMENT**

T.T.M. thanks the FWO Flanders for the financial support under project G.0260.12. H.G.T.L. thanks the Vietnamese Government and KU Leuven for a doctoral fellowship. T.N.P.V. thanks KU Leuven for the financial support (START1/09/028). The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Hercules Foundation and the Flemish Government – department EWI. The calculations were performed on the HPC cluster Tier-2 of KU Leuven.

## REFERENCES

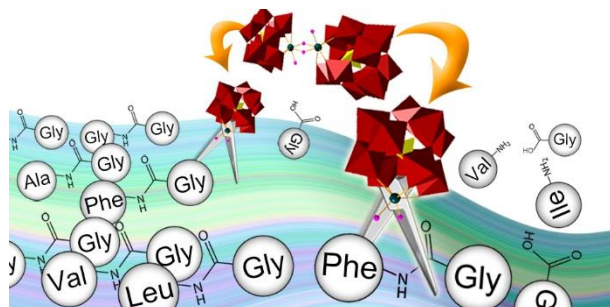
- (a) Aebersold, R.; Mann, M. *Nature* **2003**, *422*, 198-207; (b) Gallop, M. A.; Barrett, R. W.; Dower, W. J.; Fodor, S. P. A.; Gordon, E. M. *J. Med. Chem.* **1994**, *37*, 1233-1251; (c) Cruz-Migoni, A.; Fuentes-Fernandez, N.; Rabbitts, T. H. *Medchemcomm* **2013**, *4*, 1218-1221.
- Radzicka, A.; Wolfenden, R. *J. Am. Chem. Soc.* **1996**, *118*, 6105-6109.
- (a) Kozhevnikov, I. V. *Chem. Rev.* **1998**, *98*, 171-198; (b) Kozhevnikov, I. V., *Catalysis by Polyoxometalates*. John Wiley and Sons, LTD: Chichester, 2002; Vol. 2; (c) Izarova, N. V.; Pope, M. T.; Kortz, U. *Angew. Chem. Int. Ed.* **2012**, *51*, 9492-9510; (d) Sartorel, A.; Bonchio, M.; Campagna, S.; Scandola, F. *Chem. Soc. Rev.* **2013**, *42*, 2262-2280.
- (a) Muller, A.; Peters, F.; Pope, M. T.; Gatteschi, D. *Chem. Rev.* **1998**, *98*, 239-272; (b) Yamase, T. *Chem. Rev.* **1998**, *98*, 307-326.
- (a) Stephan, H.; Kubeil, M.; Emmerling, F.; Muller, C. E. *Eur. J. Inorg. Chem.* **2013**, *2013*, 1585-1594; (b) Rhule, J. T.; Hill, C. L.; Judd, D. A. *Chem. Rev.* **1998**, *98*, 327-357.
- Mizuno, N.; Misono, M. *Chem. Rev.* **1998**, *98*, 199-218.
- Sadakane, M.; Steckhan, E. *Chem. Rev.* **1998**, *98*, 219-237.
- (a) Lv, H.; Geletii, Y. V.; Zhao, C.; Vickers, J. W.; Zhu, G.; Luo, Z.; Song, J.; Lian, T.; Musaev, D. G.; Hill, C. L. *Chem. Soc. Rev.* **2012**, *41*, 7572-7589; (b) Sumliner, J. M.; Lv, H.; Fielden, J.; Geletii, Y. V.; Hill, C. L. *Eur. J. Inorg. Chem.* **2014**, *2014*, 635-644.
- (a) Ishimoto, R.; Kamata, K.; Mizuno, N. *Angew. Chem. Int. Ed.* **2012**, *51*, 4662-4665; (b) Kamata, K.; Yonehara, K.; Sumida, Y.; Yamaguchi, K.; Hikichi, S.; Mizuno, N. *Science* **2003**, *300*, 964-966.
- (a) Ly, H. G. T.; Absillis, G.; Bajpe, S. R.; Martens, J. A.; Parac-Vogt, T. N. *Eur. J. Inorg. Chem.* **2013**, *2013*, 4601-4611; (b) Ly, H. G. T.; Absillis, G.; Parac-Vogt, T. N. *Dalton Trans.* **2013**, *42*, 10929-10938; (c) Ly, H. G. T.; Absillis, G.; Janssens, R.; Proost, P.; Parac-Vogt, T. N. *Angew. Chem. Int. Ed.* **2015**, *54*, 7391-7394.
- (a) Kato, C. N.; Negishi, S.; Yoshida, K.; Hayashi, K.; Nomiya, K. *Appl. Catal., A* **2005**, *292*, 97-104; (b) Kholdeeva, O. A.; Maksimovskaya, R. I. *J. Mol. Catal. A: Chem.* **2007**, *262*, 7-24; (c) Kholdeeva, O. A.; Maksimov, G. M.; Maksimovskaya, R. I.; Vanina, M. P.; Trubitsina, T. A.; Naumov, D. Y.; Kolesov, B. A.; Antonova, N. S.; Carbo, J. J.; Poblet, J. M., *Inorg. Chem.* **2006**, *45*, 7224-7234.
- (a) Mizuno, N.; Yamaguchi, K.; Kamata, K. *Coord. Chem. Rev.* **2005**, *249*, 1944-1956; (b) Mizuno, N.; Kamata, K.; Yamaguchi, K. *Top. Catal.* **2010**, *53*, 876-893; (c) Kato, C. N.; Hayashi, K.; Negishi, S.; Nomiya, K. *J. Mol. Catal. A: Chem.* **2007**, *262*, 25-29; (d) Estrada, A. C.; Santos, I. C. M. S.; Simoes, M. M. Q.; Neves, M. G. P. M. S.; Cavaleiro, J. A. S.; Cavaleiro, A. M. V. *Appl. Catal., A* **2011**, *392*, 28-35.
- Kikukawa, Y.; Yamaguchi, S.; Nakagawa, Y.; Uehara, K.; Uchida, S.; Yamaguchi, K.; Mizuno, N. *J. Am. Chem. Soc.* **2008**, *130*, 15872-15878.
- Boglio, C.; Lemiere, G.; Hasenknopf, B.; Thorimbert, S.; Lacote, E.; Malacria, M., *Angew. Chem. Int. Ed.* **2006**, *45*, 3324-3327.
- Sartorel, A.; Carraro, M.; Scorrano, G.; Bassil, B. S.; Dickman, M. H.; Keita, B.; Nadjo, L.; Kortz, U.; Bonchio, M. *Chem. Eur. J.* **2009**, *15*, 7854-7858.
- Yasuda, H.; He, L. N.; Sakakura, T.; Hu, C. W. *J. Catal.* **2005**, *233*, 119-122.
- Zhao, S.; Huang, L.; Song, Y.-F. *Eur. J. Inorg. Chem.* **2013**, *2013*, 1659-1663.
- Orlandi, M.; Argazzi, R.; Sartorel, A.; Carraro, M.; Scorrano, G.; Bonchio, M.; Scandola, F. *Chem. Commun.* **2010**, *46*, 3152-3154.

19. (a) Absillis, G.; Parac-Vogt, T. N. *Inorg. Chem.* **2012**, *51*, 9902-9910; (b) Vanhaecht, S.; Absillis, G.; Parac-Vogt, T. N. *Dalton Trans.* **2013**, *42*, 15437-15446; (c) Ly, H. G. T.; Absillis, G.; Parac-Vogt, T. N. *New J. Chem.* **2016**, *40*, 976-984.
20. (a) Ly, H. G. T.; Absillis, G.; Parac-Vogt, T. N. *Eur. J. Inorg. Chem.* **2015**, *2015*, 2206-2215; (b) Stroobants, K.; Absillis, G.; Shestakova, P. S.; Willem, R.; Parac-Vogt, T. N. *J. Cluster Sci.* **2014**, *25*, 855-866.
21. (a) Stroobants, K.; Moelants, E.; Ly, H. G. T.; Proost, P.; Bartik, K.; Parac-Vogt, T. N., *Chem. Eur. J.* **2013**, *19*, 2848-2858; (b) Stroobants, K.; Goovaerts, V.; Absillis, G.; Bruylants, G.; Moelants, E.; Proost, P.; Parac-Vogt, T. N. *Chem. Eur. J.* **2014**, *20*, 9567-9577; (c) Stroobants, K.; Absillis, G.; Moelants, E.; Proost, P.; Parac-Vogt, T. N. *Chem. Eur. J.* **2014**, *20*, 3894-3897.
22. Luong, T. K. N.; Shestakova, P.; Mihaylov, T. T.; Absillis, G.; Pierloot, K.; Parac-Vogt, T. N. *Chem. Eur. J.* **2015**, *21*, 4428-4439.
23. Ly, H. G. T.; Mihaylov, T.; Absillis, G.; Pierloot, K.; Parac-Vogt, T. N. *Inorg. Chem.* **2015**, *54*, 11477-11492.
24. (a) Kato, C. N.; Shinohara, A.; Hayashi, K.; Nomiya, K. *Inorg. Chem.* **2006**, *45*, 8108-8119; (b) Nomiya, K.; Saku, Y.; Yamada, S.; Takahashi, W.; Sekiya, H.; Shinohara, A.; Ishimaru, M.; Sakai, Y. *Dalton Trans.* **2009**, 5504-5511.
25. Glasoe, P. K.; Long, F. A., *J. Phys. Chem.* **1960**, *64*, 188-190.
26. Neese, F. *WIREs Comput. Mol. Sci.* **2012**, *2*, 73-78.
27. (a) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098-3100; (b) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822-8824.
28. Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297-3305.
29. Andrae, D.; Haeussermann, U.; Dolg, M.; Stoll, H.; Preuss, H. *Theor. Chim. Acta* **1990**, *77*, 123-141.
30. Neese, F. *J. Comp. Chem.* **2003**, *24*, 1740-1747.
31. (a) Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *240*, 283-290; (b) Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. *Theor. Chem. Acc.* **1997**, *97*, 119-124.
32. Weigend, F. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057-1065.
33. Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, *5*, 799-805.
34. (a) Yan, L.; López, X.; Carbó, J.; Sniatynsky, R.; Duncan, D.; Poblet, J. *J. Am. Chem. Soc.* **2008**, *130*, 8223-8233; (b) Mal, S. S.; Tröppner, O.; Ivanović-Burmazović, I.; Burger, P. *Eur. J. Inorg. Chem.* **2013**, *2013*, 1960-1967; (c) López, X.; de Graaf, C.; Maestre, J. M.; Bénard, M.; Rohmer, M.-M.; Bo, C.; Poblet, J. M. *J. Chem. Theor. Comput.* **2005**, *1*, 856-861; (d) López, X.; Fernández, J.; Romo, S.; Paul, J.; Kazansky, L.; Poblet, J. *J. Comput. Chem.* **2004**, *25*, 1542-1549; (e) Fernandez, J.; López, X.; Bo, C.; de Graaf, C.; Baerends, E.; Poblet, J. *J. Am. Chem. Soc.* **2007**, *129*, 12244-12253; (f) Fernández, J. A.; López, X.; Poblet, J. M. *J. Mol. Catal. A: Chem.* **2007**, *262*, 236-242.
35. (a) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648-5652; (b) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785-789; (c) Stevens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frish, M. J. *J. Phys. Chem.* **1994**, *98*, 11623-11627.
36. Neese, F.; Wennmohs, F.; Hansen, A.; Becker, U. *Chem. Phys.* **2009**, *356*, 98-109.
37. (a) Grimme, S.; Ehrlich, S.; Goerigk, L. *J. Comput. Chem.* **2011**, *32*, 1456-1465; (b) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104-19.
38. , Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li,

- X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A. Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J., Gaussian 09, Revision D.01, *Gaussian, Inc.*, Wallingford CT, **2013**.
39. Zhao, Y.; Truhlar, D. G. *Theor. Chem. Account* **2008**, *120*, 215-241.
40. Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 299-310.
41. (a) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995-2001; (b) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669-681.
42. Boese, A. D.; Martin, J. M. L. *J. Chem. Phys.* **2004**, *121*, 3405-16.
43. (a) Pliego, J. R., Jr.; Riveros, J. M. *J. Phys. Chem. A* **2001**, *105*, 7241-7247; (b) Bryantsev, V. S.; Diallo, M. S.; Goddard III, W. A. *J. Phys. Chem. B* **2008**, *112*, 9709-9719.
44. Zamyatnin, A. A. *Prog. Biophys. Mol. Biol.* **1972**, *24*, 107-123.
45. Constantino, E.; Rimola, A.; Rodríguez-Santiago, L.; Sodupe, M. *New J. Chem.* **2005**, *29*, 1585-1593.
46. (a) Bassil, B. S.; Dickman, M. H.; Kortz, U. *Inorg. Chem.* **2006**, *45*, 2394-2396; (b) Fang, X.; Anderson, T. M.; Hill, C. L. *Angew. Chem. Int. Ed.* **2005**, *44*, 3540-3544; (c) Gaunt, A. J.; May, I.; Collison, D.; Helliwell, M. *Acta Cryst. Sec. C* **2003**, *59*, i65-i66; (d) Gaunt, A. J.; May, I.; Collison, D.; Holman, K. T.; Pope, M. T. *J. Mol. Struct.* **2003**, *656*, 101-106; (e) Villanneau, R.; Carabineiro, H.; Carrier, X.; Thouvenot, R.; Herson, P.; Lemos, F.; Ribeiro, F. R.; Che, M. *J. Phys. Chem. B* **2004**, *108*, 12465-12471; (f) Carabineiro, H.; Villanneau, R.; Carrier, X.; Herson, P.; Lemos, F.; Ramôa Ribeiro, F.; Proust, A.; Che, M. *Inorg. Chem.* **2006**, *45*, 1915-1923.
47. Jiménez-Lozano, P.; Carbó, J. J.; Chaumont, A.; Poblet, J. M.; Rodríguez-Fortea, A.; Wipff, G. *Inorg. Chem.* **2014**, *53*, 778-786.
48. Boglio, C.; Hasenknopf, B.; Lenoble, G.; Rémy, P.; Gouzerh, P.; Thorimbert, S.; Lacôte, E.; Malacria, M.; Thouvenot, R. *Chem. Eur. J.* **2008**, *14*, 1532-1540.
49. Ho, J.; Coote, M. L. *Theor. Chem. Acc.* **2009**, *125*, 3-21.
50. Burgess, J. *Metal Ions in Solution*. Halsted Press (Wiley): New York, **1978**.
51. (a) Kassai, M.; Ravi, R. G.; Shealy, S. J.; Grant, K. B. *Inorg. Chem.* **2004**, *43*, 6130-6132; (b) Kassai, M.; Grant, K. B. *Inorg. Chem. Commun.* **2008**, *11*, 521-525.

For Table of Contents Only

Graphic



Synopsis

Two is more than one! Dimeric Zr(IV)-substituted Keggin-type polyoxometalate dissociate into monomers to promote peptide bond hydrolysis in dipeptides with aliphatic side chains. The catalytic effect is due to electrophilic activation of the peptide carbonyl group by the Zr(IV) ion while the selectivity is controlled by the  $\beta$ -branching of the aliphatic side chain through intramolecular steric hindrance in the rate-limiting transition state.